

HOZZÁFÉRÉS ÉS ÚJRAHASZNOSÍTÁS: A TEL ÉS A EUROPEANA A TUDOMÁNYOS KUTATÁS SZOLGÁLATÁBAN

KOVÁCSNÉ KORENY ÁGNES

Fővárosi Szabó Ervin Könyvtár, főigazgató-helyettes

ABSZTRAKT

The European Library (TEL) és Europeana – két sokat hallott és emlegetett név, két sokszor hivatkozott szolgáltatás, sok hasonlósággal és sok különbséggel. A TEL felületén mintegy 200 millió forrás várja, hogy felfedezzék: 24 millió oldalnyi teljes szövegű dokumentum, 18 millió digitális objektum és 119 millió bibliográfiai rekord Európa 48 országának nemzeti, illetve legfontosabb egyetemi és szakkönyvtáraiból. Új platformja olyan új keresési lehetőségekkel teszi elérhetővé a rekordokat, a digitális objektumokat és a dokumentumokat, amelyekkel a találatok elemzése is minden eddigénél könnyebb. Az objektumok szakterületi felosztásával, a CERIF tárgyszavak használatával azonos tárgykörű vagy időhöz (dátumhoz) kapcsolható keresések indíthatók; a találatok pedig egyszerűen exportálhatók a legismertebb hivatkozás- és forráskezelő rendszerekhez. A Europeana 30 különböző nyelven működő, személyre szabható felületéről több mint 39 millió digitális dokumentum érhető el, melyek a kontinens legjelentősebb közgyűjteményeiből, könyvtárakból, múzeumokból, levéltárakból származnak, s közel 50 százalékuk szabadon vagy feltételekkel, de újrahasznosítható. Az előadás bemutatja a két, európai uniós támogatásból megvalósult portált, rövid fejlődéstörténetüket, az elmúlt évek fejlesztéseinek eredményeit, s a tudományos kutatás számára kifejlesztett szolgáltatásaikat, különös tekintettel arra, hogy hogyan teszik könnyen hozzáférhetővé és mélységeiben is kutathatóvá ezt a hatalmas, soknyelvű, rendkívül heterogén adathalmazt a nyílt hozzáférésű adatok összekapcsolása révén.

Tudományos kutatók a virtuális térben

A *Centre for Comparative and International Business Research* (CIBER) 2012-ben tette közzé egy közel tíz évig tartó kutatás eredményeit, amelynek során a virtuális térben kutatók információkeresési és olvasási szokásait vizsgálták. Az eredmények azt mutatták, hogy a virtuális környezetben a kutatók rendkívül aktívak, de a munka nagy részét a keresőrobotokkal végeztetik. Közvetlen hozzáférést szeretnének a forrásokhoz, s lehetőleg digitálisan, drót nélkül, a saját eszközükön. Az interneten szeretnek „ugrálni”, minél többet, innen oda: a több ezer elérhető oldalból egyet vagy kettőt néznek meg, s 40%-uk nem megy vissza többé az egyszer már megtekintett oldalhoz. Az „ugrálás” oka, hogy a kutatáshoz a legismertebb keresőmotorokat használják, amelyek (túl) sok (nem releváns) találatot és rengeteg linket ajánlanak fel számukra; gyenge a keresési kompetenciájuk, s kevés az idejük.

A kutatók a virtuális térben másképp, s gyorsan olvasnak: a horizontális irányú szövegpáštázás felváltotta a vertikális, soronkénti olvasást, a szövegben a kulcsszavakat

keresik, s gyakran csak a címet, a heading-et és összefoglalót (abstract) futják át, majd mennek is tovább. Néhány percet töltenek csak a kereséssel, illetve egy-egy oldalon; a valódi olvasás helyett csak rápillantanak a találatokra. Az érdekesebbnek tűnő cikkekre 3–4 percet szánnak, s inkább elmentik, letöltik maguknak későbbi használatra.

Általában nem is használják a szépen kialakított, rájuk szabott keresőfelületeket, inkább „hátról”, a linkeken keresztül érkeznek meg egy-egy releváns találathoz. S hogy számukra mi releváns, azt sokkal inkább befolyásolja az, hogy társaik, barátaik mit ajánlanak a közösségi oldalakon, mint az, hogy egy publikáció mely nagy múltú lektorált folyóiratban jelent meg.

A tudományos kutatás módszereit, illetve a kutatók információ-szerzési tapasztalatait vizsgáló tanulmányok, s az azokból jól leírható viselkedési, információgyűjtési, keresési és feldolgozási minták adták az alapját a TEL és a Europeana elmúlt években végrehajtott fejlesztéseinek.

A TEL és a Europeana rövid fejlődéstörténete, hasonlóságai és főbb különbségei

A TEL története 1997-ben, a GABRIEL projekttel indult, melynek célja az volt, hogy egy közös portál jöjjön létre Európa nemzeti könyvtárai és katalógusaik számára. Az új oldal 2005 márciusában kezdte meg működését. Még ugyanebben az évben az Európai Parlament kezdeményezte egy európai szintű közgyűjteményi digitális portál, a Europeana felállítását, amelyet 2008 novemberében meg is nyitottak a felhasználók számára.

Noha mindkét szolgáltatás központja a Holland Nemzeti Könyvtár (Koninklijke Bibliotheek) Hágában, üzemeltetőik ma már mások: a TEL-t az Európai Nemzeti Könyvtárak Konferenciája (Conference of European National Librarians = CENL), az Európai Szakkönyvtárak Szövetsége (Ligue des Bibliothèques Européennes de Recherche = Association of European Research Libraries = LIBER) és az Európai Szakkönyvtárak Konzorciuma (Consortium of European Research Libraries = CERL) működteti, a Europeana-t pedig a Europeana Foundation.

A *The European Library* a Europeana Libraries projekt lezárása, 2012 vége óta „A” könyvtári aggregátor a Europeana számára, mintegy 48 nemzeti könyvtár és a vezető európai szakkönyvtárak bibliográfiai rekordjaihoz és digitális objektumaihoz, összesen több mint 28,5 millió digitális objektumhoz és 159,3 millió bibliográfiai rekordhoz kínálva hozzáférést. A *Europeana* az európai közgyűjtemények digitális táráként vezető európai könyvtárak, múzeumok, levéltárak és archívumok digitális forrásait teszi elérhetővé. Ez számszerűleg több mint 52,2 millió digitális objektumot, köztük 27,8 millió képet, 22,3 millió szöveget, 1,1 millió videót, több mint 800 ezer hangzó anyagot és 25 ezer 3D-s dokumentumot jelent.¹

A két szolgáltatás célja közös: lehetővé tenni a felhasználók számára az információkeresést és a tudáshoz való hozzáférést Európa valamennyi nyelvén – közvetlenül, a portálokon keresztül és közvetetten, az adatokra épített különböző alkalmazások révén.

A kihívás

A célok mellett közös a kihívás is: a rendkívül heterogén adatok integrációja, azaz hogy hogyan lehet a kor követelményeinek és a felhasználók elvárásainak megfelelő, homogén tárgyszó alapú hozzáférést, soknyelvű keresést, illetve a leírt források közötti szemantikai kapcsolatokra is építő szolgáltatásokat biztosítani.

A megoldás a tudásszervezési rendszerek (Knowledge Organisation Systems = KOS) alkalmazása:

- a Europeana Libraries projekt keretében kialakított, saját fejlesztésű infrastruktúra és munkafolyamat az adatbevitelre, az indexálásra, az adatnormalizációra és az adatok közzétételére;
- a Europeana Data Model (EDM) metaadat szabványt jelent, amely a szemantikus web elveit, megoldásait kihasználva képes a legkülönbözőbb adatok nyílt környezetben való integrálására és újrahasznosítására. Az EDM adatstruktúra a szemantikus web alap gondolatát (vagyis, hogy miként kapcsoljunk metaadatokat, metainformációkat az interneten elérhető forrásokhoz, gazdagítva így egyrészt az adatokat, másrészt lehetővé téve, hogy az adatokat az automatizált eszközök is képesek legyenek egymással megosztani és feldolgozni) és megoldásait (a különböző névtér adatbázisok és tudásszervezési rendszerek integrációja révén megvalósuló adatgazdagítás lehetőségeit) kihasználva képes a legkülönbözőbb adatok nyílt környezetben való integrálására és újrahasznosítására.

A megoldás

A tudásszervezési rendszerek a dokumentumok és fogalmak összerendezését, rendszerré szervezését végzik; plusz „tartalommal” gazdagítják az objektumokat. Ilyenek a soknyelvű szótárak, a teauruszok, az authority listák, a névterek, az osztályozási rendszerek, az ontológiák vagy a taxonómiák.

A TEL és a Europeana olyan tudásszervezési rendszereket keresett, amelyek műszakilag megfelelők és elérhetők (Linked Data-n keresztül vagy dedikált repozitóriumban), jól dokumentáltak, jól rendezettek, open access felhasználásúak és soknyelvűek. Alkalmazásuk a két szolgáltatás esetében szinte azonos.

The European Library

A TEL legnagyobb problémája, hogy óriási mennyiségű besorolási adattal rendelkezik a különböző európai könyvtáraktól aggregált bibliográfiai adatoknak köszönhetően. A rekordkészítésnél a könyvtárak sokféle rendszert, sokféle gyakorlatot alkalmaznak, s természetesen sokféle nyelven készülnek a rekordok. Súlyosbítja a helyzetet, hogy a könyvtárak többsége helyi, nyelvfüggő rendszereket használ a szerzőségi adatok és a tartalom leírására. A legfontosabb feladatot tehát a szerzőségi adatok konszolidációja és a tárgyszó-normalizálás jelenti a TEL számára.

A TEL a megoldást a különböző nemzetközi és nyelvfüggetlen rendszerek használatában találta meg, melyek segítségével 2012 végére 120 000 linket sikerült bekötniük

a francia, brit, német és svájci nemzeti könyvtárakból származó tárgyszavak közé. Minden bekerülő bibliográfiai rekord tárgyszó- és névadat normalizáció alá esik, így mára sikerült egy, a nyelveket „átszövő”, vagyis az egyes nyelvek közötti kétirányú átjárást lehetővé tevő tárgyszóbázist kialakítani.

A TEL által használt tudásszervezési rendszerek a következők:

- Nemzetközi Virtuális Azonosító Adattár (Virtual International Authority File = VIAF), honlapja: <http://www.viaf.org>;
- a Nemzetközi Szabványos Névazonosító (International Standard Name Identifier = ISNI), honlapja: <http://www.isni.org>;
- a Közös Európai Kutatási Információs Formátum (Common European Research Information Format = CERIF) részeként kialakított Közös Európai Kutatási Osztályozási Rendszer (Common European Research Classification Scheme);
- a Kongresszusi Könyvtár Tárgyszórendszere (Library of Congress Subject Headings = LCSH); honlapja: <http://id.loc.gov/authorities/subjects.html>
- a Francia Nemzeti Könyvtár Tárgyszórendszere (Répertoire d'autorité-matière encyclopédique et alphabétique unifié = RAMEAU), honlapja: [http://\[-\]rameau.bnf.fr](http://[-]rameau.bnf.fr)
- a Német Nemzeti Könyvtár tárgyszórendszere (Schlagwortnormdatei = SWD).

Illetve ezen tárgyszórendszerek kapcsolatrendszere, az ún. MACS (Multilingual Access to Subjects = Soknyelvű Hozzáférés a Tárgyszavakhoz), amely program az angol, német és francia tárgyszórendszereket kapcsolja össze és felelteti meg egymással. (Honlapja: http://www.dnb.de/EN/Wir/Kooperation/MACS/macs_node.html.)



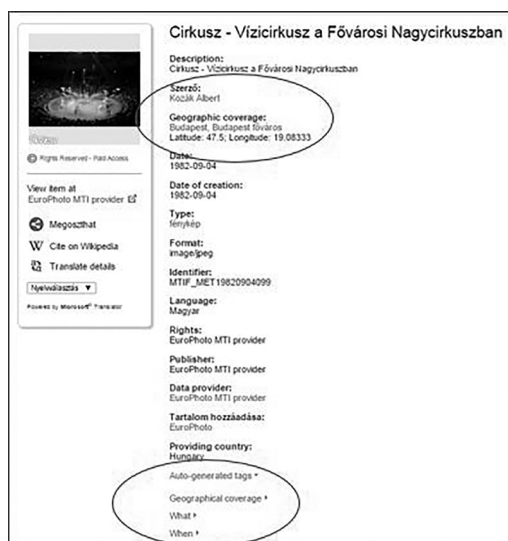
1. ábra: Adatgazdagítás a Europeanában

Europeana

A gyűjtött források rendkívüli különbözősége, sokszínűsége miatt a korábban jelzett problémák megoldására a Europeana az automatikus adatgazdagítás lehetőségeit igyekszik kihasználni. Az alább felsorolt külső hivatkozási források segítségével a *ki*, *mit*, *hol* és *mikor* kérdésekre adnak választ, ezzel pontosítva a keresést, s növelve a releváns találatok számát a soknyelvű, rendkívül heterogén adathalmazban.

A Europeana a következő rendszereket alkalmazza az adatgazdagításra (lásd 1. ábra):

- A *hol?* kérdésre a GeoNames ontológiát veszi igénybe, amely egy, a világhálón szabadon hozzáférhető, több mint 8 millió földrajzi nevet tartalmazó névtér. A Europeana jelenleg több mint 7 millió digitális objektuma van hozzárendelve valamely, a GeoNames-ben rögzített földrajzi névhez. A GeoNames honlapja: <http://www.geonames.org>.
- A *ki?* kérdésre a DBpedia adatbázist hívja segítségül, amely a Wikipédia adatait tartalmazza rendezett, strukturált formában. Ez jelenleg (2016 elején) több mint 4,58 millió bejegyzést jelent. Ebből 1 445 000 személy, 735 000 hely, 123 000 zenei album, 87 000 film, 19 000 videojáték, 241 000 szervezet, 251 000 faj és 6 000 betegség. Az adatok 125 különböző nyelven érhetők el. Az adatbázis 25,2 millió képhivatkozást és 29,8 millió külső hivatkozást is tartalmaz.² Honlapja: <http://wiki.dbpedia.org>.
- A *mi?* kérdés esetében a GEMET (GEneral Multilingual Environmental Thesaurus) nevű általános, többnyelvű környezetvédelmi tezausz és a DBpedia bővíti a válaszokat.
- A *mikor?* kérdésre – több mint 10 millió objektum esetében – a Semium Time alkalmazás segítségével ad választ (lásd 2. ábra).



2. ábra: Az idővel kapcsolatos adatok gazdagítása a Europeana-ban

További szolgáltatások

A szemantikus web adta lehetőségek alkalmazásán túl a két portál számos speciális szolgáltatást kínál az érdeklődőknek és kutatóknak egyaránt.

The European Library (<http://www.theeuropeanlibrary.org>)

A TEL portál az elmúlt években nagyszabású fejlesztésen ment keresztül, s új külsőt is kapott, hogy jobban megfeleljen a kutatók elvárásainak és a Europeana kívánalmainak. Az új felület nagy mennyiségű bibliográfiai rekordot, digitális objektumot és teljes szövegű dokumentumot tesz elérhetővé a felhasználók számára, s olyan új keresési lehetőségeket, amelyekkel a találatok elemzése is könnyebbé válik:

- Egyszerű tárgyszavas keresés és böngészés, a találatok elemzése, szűkítése idő, hely és tárgy szerint.
- Háromféle Open Search API (Application Programming Interface = alkalmazás-programozási felület, felhasználói program-interfész), vagyis egy nyílt hozzáférésű kereső program API³-regisztrált felhasználóknak. Az Open Search API segítségével a TEL szolgáltatásában automatizálható a keresés, hiszen az a szolgáltató indexei alapján dolgozik.
 - o The European Library OpenSearch API: a TEL portálon elérhető, a nemzeti és szakkönyvtárak gyűjteményeit leíró adatokhoz való hozzáférés.
 - o The European Library Open Dataset: hozzáférés több mint 82 millió bibliográfiai rekordhoz a Creative Commons CC0 1.0 Universal licenc alapján.
 - o Research Libraries UK (RLUK) Database: közel 20 millió bibliográfiai rekord az Egyesült Királyság 34 könyvtárából.
- Szabványos letöltési formátumok sorozata, lehetővé téve a rekordok közvetlen exportját a legnépszerűbb hivatkozási szoftverekbe, mint például a Mendeley és a Zotero.

Europeana (<http://www.europeana.eu>)

A Europeana 48,5 millió műalkotást, műtárgyat, könyvet, videót és hanganyagot kínál a kutatóknak és az érdeklődőknek. Új, tematikus gyűjteményeket alakítottak ki a zene és a művészettörténet területén, a virtuális kiállítások egy-egy aktuális téma dokumentum-állományába adnak betekintést. A megszokott keresési lehetőségek mellett a Europeana speciális, kutatókra szabott szolgáltatásokat is nyújt:

- Egyszerű tárgyszavas keresés és böngészés, a találatok elemzése, szűkítése idő, hely és tárgy szerint az általános érdeklődésű felhasználók számára.
- Europeana Research projekt és portál a kutatás és a kutatók támogatására (<http://research.europeana.eu>).
- Open Search API-k a regisztrált felhasználóknak (<http://labs.europeana.eu/api>) (a regisztráció egy egyszerű, név, intézmény, e-mail cím mezőket tartalmazó űrlap kitöltését jelenti):

- o REST-API – az adatok dinamikus visszakeresésére (REST = Representational State Transfer, azaz reprezentációs állapot átvitel architektúra, amely az adatok kéréskészleten keresztül elérését teszi lehetővé⁴).
- o OAI-PMH – metaadatgyűjtő protokoll az adatok aratására (Open Archives Initiative – Protocol for Metadata Harvesting, azaz Nyílt Archívumok Kezdeményezés a metaadatok aratására, amelynek célja a digitális archívumokban található információk hozzáférhetőségének, ismertségének növelése).
- Speciális kutatásokat lehetővé tevő fejlesztéseket is találunk a Europeana Research projektben, többek között:
 - o Hogyan lehet a Europeana Newspapers projekt teljes szövegeit és képeit újrafelhasználásra elérhetővé tenni?
 - o A Europeana Newspapers-re alapuló full-text API kifejlesztése.
 - o További gyűjteményrészek kiemelése a kutatók számára.
 - o Kutatói elvárások, igények és szükségletek folyamatos feltérképezése.
 - o Tartalomfejlesztési stratégia kidolgozása a kutatói igények alapján.
 - o Ösztöndíjprogram kidolgozása kutatók számára.

Összegzésként elmondhatjuk, hogy a szemantikus web kínálja megoldások folyamatosan fejlődő alkalmazásával a TEL és a Europeana hatékonyan biztosítja, hogy az európai közgyűjtemények kulturális kincsei digitális formában, a lehető leghatékonyabban kereshető és elemezhető módon elérhetővé váljanak a tudományos kutatók számára. Mégpedig úgy, ahogy azt ők szeretnék.

Irodalom

- CHARLES, Valentine – FREIRE, Nuno – ISAAC, Antoine: Links, languages and semantics: linked data approaches in The European Library and Europeana. Forrás: https://www.researchgate.net/profile/Nuno_Freire/publication/271907664_Links_languages_and_semantics_linked_data_approaches_in_The_European_Library_and_Europeana/links/54d61fa30cf2970e4e67d21d.pdf [2016. január 7.]
- FREIRE, Nuno: Facilitating Access and Reuse of Research Materials: the Case of The European Library. Forrás: http://elpub.scix.net/data/works/att/107_elpub2013.content.08073.pdf [2016. január 7.]
- FREIRE, Nuno [et al.]: Author Consolidation across European National Bibliographies and Academic Digital Repositories. Forrás: http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_22_full_paper.pdf [2016. január 7.]
- ISAAC, Antoine: Classification schemes, thesauri and other Knowledge Organisation Systems – a Linked Data perspective. Forrás: <http://www.slideshare.net/antoineisaac/linked-pasts> [2016. január 7.]
- KOVÁCSNÉ KORENY Ágnes: A TEL-szolgáltatás jövője – javaslat. = Tudományos és Műszaki Tájékoztatás, 59. évf. 7. sz. 2012. 293–297. p.
- KOVÁCSNÉ KORENY Ágnes: Europeana Libraries – egy új európai projekt. = Tudományos és Műszaki Tájékoztatás, 58. évf. 7. sz. 2011. 283–289. p.

KOVÁCSNÉ KORENY ÁGNES

KOVÁCSNÉ KORENY Ágnes: The European Library. Jelenlegi és tervezett szolgáltatások könyvtáraknak és kutatóknak. = Tudományos és Műszaki Tájékoztatás, 59. évf. 7. sz. 2012. 297–298. p.

KOVÁCSNÉ KORENY Ágnes: TEL szabványosítási kézikönyv. = Tudományos és Műszaki Tájékoztatás, 59. évf. 7. sz. 2012. 298–306. p.

NICHOLAS, David - CLARK, David: Information Seeking and Reading Behaviour of the Virtual Researcher. Forrás: http://ciber-research.eu/download/20120627-Tartu_CIBER.pdf [2016. január 7.]

Jegyzetek

¹ A számadatok 2016. április közepéről származnak.

² A számadatok forrása: <https://hu.wikipedia.org/wiki/DBpedia> [2016. január 7.]

³ Az API egy program azon eljárásainak, szolgáltatásainak és azok használatának a dokumentációja, amelyet más programok felhasználhatnak. Egy nyilvános API segítségével lehetséges egy programrendszer szolgáltatásait anélkül használni, hogy annak belső működését ismerni kellene.

⁴ Forrás: IBM Knowledge Center: [http://www.ibm.com/support/knowledgecenter/SSY\[-\]MRC_4.0.1/com.ibm.rational.reporting.overview.doc/topics/c_ovr_process_rest.\[-\]html?lang=hu](http://www.ibm.com/support/knowledgecenter/SSY[-]MRC_4.0.1/com.ibm.rational.reporting.overview.doc/topics/c_ovr_process_rest.[-]html?lang=hu) [2016. január 7.]

Kovácsné Koreny Ágnes az okleveles könyvtáros, illetve a magyar nyelv és irodalom szakos bölcsész diploma megszerzése után pályáját az MTA Könyvtárában kezdte 1993-ban. 1996–2004 között az Európai Bizottság Magyarországi Delegációjának információs menedzser munkatársa. 2004-től az Országgyűlés Hivatalában dolgozott, előbb a könyvtárban, majd a Közönségszolgálati Iroda vezetőjeként. 2014 májusától a Fővárosi Szabó Ervin Könyvtár főigazgató-helyettese. Tagja a New Library World folyóirat tanácsadó bizottságának, ugyanott szakmai lektor is. Az ELTE Irodalomtudományi Doktori Iskola Könyvtártudomány Doktori Programjának hallgatója.